The difference-in-Sargan test

This package implements the difference-in-Sargan test (sometimes called the “GMM distance” test or $C$-test) for exogeneity of one or more regressors in the context of an IV model estimated via gretl's ts1s command.

The overall strategy of this test is:

1. Estimate an IV model in which the variables in question are treated as endogenous; call this the restricted model.
2. Estimate a model in which the variables in question are added to the list of instruments; call this the unrestricted model.
3. Compute the test statistic as the Sargan test for the unrestricted model minus that for the restricted.

The standard Sargan overidentification test assesses the joint null hypothesis that an IV model is correctly specified and the instruments are valid (asymptotically uncorrelated with the error term)—see Davidson and MacKinnon (1993, section 7.8) for details. The idea behind the difference-in-Sargan test is that one can equally well test the marginal increase in the degree of overidentification that results when one or more instruments are added. The logic of the test, however, requires that the initial (restricted) model is correctly specified with valid instruments, so if the standard Sargan test rejects one should not place much faith in the difference-in-Sargan test.

Command-line use

The function available for command-line and scripting use is named `exogtest`. It takes the following arguments:

<table>
<thead>
<tr>
<th>name</th>
<th>type</th>
<th>comment</th>
<th>default value</th>
</tr>
</thead>
<tbody>
<tr>
<td>model</td>
<td>bundle</td>
<td>the initial IV model</td>
<td>-</td>
</tr>
<tr>
<td>X</td>
<td>list</td>
<td>the variable(s) to be tested</td>
<td>-</td>
</tr>
<tr>
<td>quiet</td>
<td>boolean</td>
<td>quiet operation?</td>
<td>0 (no)</td>
</tr>
</tbody>
</table>

On successful completion `exogtest` returns a 3-vector holding test statistic, degrees of freedom and $p$-value.

We illustrate by reference to the sample script included in the package, which uses data on young men’s wages from Griliches (1976). The script begins with

*Department of Economics, Wake Forest University.
include exogtest.gfn
open griliches.gdt
# TSLS, schooling (s) treated as endogenous
series exprsq = expr^2
list xlist = const age expr exprsq s
list zlist = const expr exprsq age iq med
tsls lw xlist ; zlist

Here we estimate an overidentified IV model for the log of the wage, with exogenous regressors age, experience and the square of experience and potentially endogenous regressor years of schooling (s); the excluded instruments are IQ score and mother’s education. The Sargan test statistic for this regression is 0.028, with a p-value of 0.87, so we do not reject the null of correct specification and valid instruments. But now we’d like to see if schooling can be treated as exogenous:

# test the schooling variable
matrix dstest = exogtest($model, s)
print dstest

The answer is a firm “No”: the test statistic is 23.1 with p-value $1.5 \times 10^{-6}$. Note that although the second parameter to exogtest is marked as a list, it’s acceptable to provide as argument a single series.

3 Some details

If the initial model is exactly identified then the initial Sargan test is taken to be identically zero. Obviously, however, one has no assurance regarding the validity of the instruments in this case. If the initial model is overidentified and the Sargan test rejects at the 5 percent level, a warning is printed.

If the revised (unrestricted) model has no remaining endogenous regressors then it is estimated via OLS, otherwise by TSLS.

Each Sargan statistic is calculated as the explained sum of squares from an auxiliary regression of the respective residual on the full set of instruments, divided by a measure of the error variance. To ensure a non-negative result in finite samples the same variance estimate is used in both cases, namely the MSE from the unrestricted model.

Let subscript $r$ indicate the restricted model and $u$ the unrestricted, and let ESS denote the explained sum of squares from the associated auxiliary regression. Then the test statistic is calculated as $S_u - S_r$, where

$$ S_i = \frac{ESS_i}{MSE_u} \quad i = r, u $$

The test statistic is asymptotically distributed as $\chi^2$ with degrees of freedom equal to the number of regressors under test.

4 Use in the gretl GUI

This package offers to attach to the Analysis menu in the model window after estimation of a model via TSLS: the menu item is titled “Exogeneity.” Clicking this item brings up a dialog where you can select a regressor to test. (At present in the GUI you are limited to selecting a single variable.)

5 Replication exercise

Listing 1 shows an additional test case for the package. This replicates an example shown in Baum et al. (2007, section 5). It is similar to the sample script (a wage regression in which the exogeneity
of a measure of education is to be tested) but in this case the instruments for education are weak and the difference-in-Sargan test fails to reject the null hypothesis that education is exogenous.

Example 1: Replication of example from Baum et al

Input:

```
include exogtest.gfn
open mroz87.gdt -q
# restrict the sample to members of the labor force
smpl LFP --dummy

# TSLS, educ treated as endogenous
series exper = AX
series expersq = AX^2
series educ = WE
series age = WA
series lwage = log(WW)
list xlist = const exper expersq educ
list zlist = const exper expersq age KL6 K618
tsls lwage xlist ; zlist
dstest = exogtest($model, educ)
```

Partial output from test:

```
Test for exogeneity of educ:
  chi-square(1) = 0.721676 - 0.702529 = 0.0191471 [0.8899]
```

References

